



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### **eHistology image and annotation data from the Kaufman Atlas of Mouse Development**

**Citation for published version:**

Baldock, RA & Armit, C 2017, 'eHistology image and annotation data from the Kaufman Atlas of Mouse Development', *GigaScience*. <https://doi.org/10.1093/gigascience/gix131>

**Digital Object Identifier (DOI):**

[10.1093/gigascience/gix131](https://doi.org/10.1093/gigascience/gix131)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

GigaScience

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



## Title

*eHistology image and annotation data from the Kaufman Atlas of Mouse Development*

## Authors

Richard A Baldock<sup>1</sup> and Chris Armit<sup>1</sup>

## Affiliations

1. MRC Human Genetics Unit, Institute of Genomic and Molecular Medicine, University of Edinburgh, Crewe Road, Edinburgh EH4 2XU, UK

corresponding author: Richard Baldock (Richard.Baldock@igmm.ed.ac.uk)

## Abstract

The “Atlas of Mouse Development” by Kaufman is a classic paper atlas that is the *de facto* standard for the definition of mouse embryo anatomy in the context of standard histological images. We have re-digitised the original H&E stained tissue sections used for the book at high resolution and transferred the hand-drawn annotations to digital form. We have augmented the annotations with standard ontological assignments (EMAPA anatomy) and made the data freely available via an online viewer (eHistology) and from the University of Edinburgh DataShare archive. The dataset captures and preserves the definitive anatomical knowledge of the original atlas, provides a core image set for deeper community annotation and teaching, and delivers a unique high-quality set of high-resolution histological images through mammalian development for manual and automated analysis.

## Background & Summary

The “Atlas of Mouse Development” (Kaufman 1994) is a book detailing the anatomy of mouse embryo development and stands as the definitive work in the field. The atlas is based on a lifetime of work by Kaufman who established a unique set of histological sections of about 450 mouse embryos, many of which are full serial section-series, from which he selected carefully staged samples for the histological images within the book. The combination of the histological section series and the printed book represent a unique resource and captures the current understanding of classical mouse anatomy. In taxonomic terms these physical sections are the *reference specimens* for the definition of mouse embryo anatomy and the digitised images with the associated annotations are a digital *holotype* for the definition of anatomical terms and the progression of mouse embryo development. In addition the paper atlas has given rise to the Mouse Atlas programme in Edinburgh (Richardson *et al*, 2014) and to the EMAPA mouse anatomy ontology (Hayamizu *et al*, 2013; Hayamizu *et al*, 2015). The original index for the book was used to develop the primary list of anatomical terms in the ontology, and EMAPA is now recognised as the standard mouse embryo ontology used to annotate mouse embryo data including embryo phenotype data (Brown and Moore, 2012; Mohun *et al*, 2013).

In generating the eHistology Atlas, new images of the histological sections were acquired at high resolution, and the annotations have been transferred to a database. These images and annotations are now freely available from the eMouseAtlas web resource as eHistology (see figure 1) and have been described by Graham *et al* (2015). Here we describe the dataset of the 937 high-resolution histology images with anatomy annotations and how they have been made available for further study and analysis.

The motivation for the eHistology resource was to capture the anatomical knowledge in a permanently accessible open and digital form delivered with a viewer providing a view of the underlying histology data not possible in the printed atlas. The high-resolution images

provide a rich resource of carefully staged mouse histology, which could be used for deeper analysis of tissue development and as a teaching resource. See figure 1 for an illustration of the resolution now available for these images.

All the data is available under a creative commons licence (CC BY 4.0). In the future, we envisage the annotations being extended on a tissue-by-tissue basis through community curation. The histology viewer is open-source and is available from the Mouse Atlas technical GitHub repository ([github.com/ma-tech](https://github.com/ma-tech)).

Providing for a secure (in terms of data preservation) and long-term accessibility for research data is a difficult problem. A recent study of the longevity of 375 biomedical resources/databases (Attwood et al, 2015) available on the web in 1997 found that 62.3% had ceased to be available, 14.4% were static and only 23.3% available as an active resource. The authors concluded that survival depended primarily on Institutional interest and that a strategy dependent on external funding will very likely fail. To ensure long-term preservation of the image data and supporting annotations, we have therefore registered this dataset with the University of Edinburgh *DataShare* (<http://datashare.is.ed.ac.uk/>) repository (Rice 2014) with policies registered in OpenDOAR (Directory of Open Access Repositories, <http://www.opendoar.org/find.php?rID=1176&format=full>). Specifically the preservation policy includes indefinite preservation of the original data with format migration to ensure continued readability and accessibility. We suggest that the longevity provided by the University of Edinburgh will exceed any other option not associated with a similar institution.

## Methods

### Histology

Details of the mouse strains used, histological sectioning and staining are provided by Kaufman (1994). Briefly the embryos were “isolated from spontaneously cycling (C57BL X CBA) F1 hybrid females that had been previously mated to genetically similar F1 hybrid males”. The embryos were fixed, dehydrated, embedded in paraffin wax and sectioned at seven-micron thickness. The mounted sections were then stained with haematoxylin and eosin.

### Slide digitization

Digitization of the original histology slides was accomplished using the Olympus DotSlide slide scanner system. Using a  $\times 20$  objective lens, this generated full colour images with a pixel resolution of 0.34 microns. Calibration was accomplished as part of the digitization process, allowing the inclusion of scale-bars and the option to measure the distance between two points. In two instances (Plate 5 and Plate 14) the original sections could not be sourced and were presumed lost. In these instances the original photographic negatives were used in place of the original slides to generate cellular-resolution grey-scale images.

### Annotation and linking to the EMAPA ontology

Annotation was accomplished using a manual procedure whereby “flags” were positioned on points corresponding to the matching points as used in each plate in the book. The flags were placed using a editors version of the eHistology interface (Graham *et al*, 2015). Each flag was linked to the anatomical term or phrase used in the book and also an EMAPA ontology term and an associated Wikipedia link. There were over 10,000 flag labels used to annotate the eHistology sections, and linking them to EMAPA IDs was achieved through a combination of string matching and manual assignment of terms (Richardson et al., 2015). Linking to Wikipedia was accomplished using a manual process that utilized parent terms in the *partonomic* ontology tree to find the closest match for a given anatomical term or tissue.

## Code availability

All code used to generate this dataset is available open-source from the ma-tech repository at GitHub.

## Data Records

Each record has an assigned Digital Object Identifier (DOI) that resolves to a set of data files comprising a jpeg or tiff encoded image, Dublin core and other metadata files and the set of annotations associated with the image. The image data volumes range up to 2Gb with a total volume of 118GB for the full series in compressed “zip” format. Table 1 lists the files with each dataset. Each University of Edinburgh DataShare submission requires a subset of the Dublin Core [dublincore.org] data elements to be completed and allows a further set of optional elements these are detailed in table 2. Table 3 provides a partial listing of the datasets available as an example of the data content. The full listing of all 937 images is provided in the supplementary Excel formatted data file SciDataKaufmanTable3.xlsx and corresponds to all of the histology section images of the original atlas for Plate numbers 2-41.

Table 1. Listing of the data files available with each dataset.

File	Description
<b>license.txt</b>	Licence agreement for the data deposited at DataShare providing permissions for distribution and migration as needed - CC BY 4.0.
<b>README.txt</b>	Short description of the data and data files.
<b>citation.txt</b>	How to cite use of this particular image.
<b>details.txt</b>	The text describing the embryo taken from the matching page of the printed atlas and provided in tab-delimited form for reading into a spreadsheet programme.
<b>image.tif/image.jpg</b>	Full resolution tiff or jpeg formatted image of the histological section.
<b>Image.txt</b>	Image pixel dimensions and pixel size in microns.
<b>terms.txt</b>	A tab-delimited table of the annotations for this image providing the Kaufman annotations, location in the image, annotation number, EMAPA ID and EMAPA term with synonyms in brackets.
<b>url.txt</b>	Text providing the URL for the image on the emousetlas.org web resource.

Table 2. DataShare Dublin Core elements used for the Kaufman datasets. The DCMI column provides the official Dublin Core term for the element, the Label is the heading for these data on the DataShare metadata listing.

Element	Qualifier	DCMI	Label	Input Type	Mandatory
<b>contributor</b>		contributor	Depositor	name	true
<b>contributor</b>	other	contributor	Funder	name	true
<b>creator</b>		creator	Data Creator	name	false
<b>date</b>	accessioned	date	Date Accessioned	date	true
<b>date</b>	available	date	Date Available	date	true
<b>identifier</b>	citation	identifier	Citation	citation	false
<b>identifier</b>	uri	identifier	Persistent Identifier	DOI/handle	true
<b>description</b>	abstract	decription	Data Description (abstract)	text	false
<b>description</b>	tableofcontents	decription	Data Description (TOC)	text	false
<b>publisher</b>		publisher	Publisher	text	true
<b>relation</b>	isversionof	isVersionOf	Relation (Is Version Of)	text	false
<b>relation</b>	isreferencedby	isReferencedBy	Relation (Is Referenced By)	text	false
<b>subject</b>		subject	Subject Keywords	text	false
<b>subject</b>	classification	subject	Subject Classification	Controlled text	false
<b>title</b>		title	Title	text	true
<b>title</b>	alternative	title	Alternative Title	text	false
<b>type</b>		type	Type	Controlled text	true

Table 3. Partial list of data records for the Kaufman Atlas image set. The DOI resolves to a data set of image data, metadata and annotations, which can be downloaded individually or combined. Image volumes range up to about 2Gb with a total volume for the full set of compressed zip files 118Gb. The Position column gives an estimate of the relative distance through the embryo of the individual histology section. The values are between 0 (zero) and 1 (one) corresponding to the proportionate distance left-right (sagittal sections), cranial-caudal (transverse sections) and dorsal-ventral (coronal sections).

Kaufman Image	Age	Stage	Orientation	Position	Stain	DOI
Plate 02 image a	E5.5	7-8	Sagittal		H&E	<a href="http://dx.doi.org/10.7488/ds/393">http://dx.doi.org/10.7488/ds/393</a>
Plate 02 image b	E5.5	7-8	Sagittal		H&E	<a href="http://dx.doi.org/10.7488/ds/394">http://dx.doi.org/10.7488/ds/394</a>
Plate 02 image c	E5.5	7-8	Sagittal		H&E	<a href="http://dx.doi.org/10.7488/ds/395">http://dx.doi.org/10.7488/ds/395</a>
Plate 02 image d	E5.5	7-8	Sagittal		H&E	<a href="http://dx.doi.org/10.7488/ds/396">http://dx.doi.org/10.7488/ds/396</a>
Plate 03 image a	E6.5,7	9,10	Transverse	0.210	H&E	<a href="http://dx.doi.org/10.7488/ds/397">http://dx.doi.org/10.7488/ds/397</a>
Plate 03 image b	E6.5,7	9,10	Transverse	0.370	H&E	<a href="http://dx.doi.org/10.7488/ds/398">http://dx.doi.org/10.7488/ds/398</a>
Plate 03 image c	E6.5,7	9,10	Transverse	0.590	H&E	<a href="http://dx.doi.org/10.7488/ds/399">http://dx.doi.org/10.7488/ds/399</a>
.	.	.	.	.	.	.
.	.	.	.	.	.	.
Full table in supplementary Excel file <a href="#">SciDataKaufmanTable3.xlsx</a>						
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
Plate 40l image b	E17.5	26	Transverse	0.888	H&E	<a href="http://dx.doi.org/10.7488/ds/1301">http://dx.doi.org/10.7488/ds/1301</a>
Plate 40l image c	E17.5	26	Transverse	0.903	H&E	<a href="http://dx.doi.org/10.7488/ds/1302">http://dx.doi.org/10.7488/ds/1302</a>
Plate 40l image d	E17.5	26	Transverse	0.921	H&E	<a href="http://dx.doi.org/10.7488/ds/1303">http://dx.doi.org/10.7488/ds/1303</a>
Plate 40l image e	E17.5	26	Transverse	0.948	H&E	<a href="http://dx.doi.org/10.7488/ds/1304">http://dx.doi.org/10.7488/ds/1304</a>
Plate 41 image 41	E17.5	26	Sagittal	0.496	H&E	<a href="http://dx.doi.org/10.7488/ds/1305">http://dx.doi.org/10.7488/ds/1305</a>

## Technical Validation

The Images and associated data are all validated against the published atlas which provides the detail of the genotype, defines the histological protocols, and establishes the correct staging of each embryo against the Theiler criteria. The section images used in the book are from specific tissue sections identified on the sets of microscope slides stored at the MRC Human Genetics Unit at the IGMM, University of Edinburgh. Each section was scanned digitally then checked by a second curator to ensure validity. The annotations were originally captured using optical character recognition (OCR) and the text and spelling checked by a second curator. All the end-point locations for the annotation terms have been double-checked and a series of quality control steps have meant that inspection of the whole data set has not revealed any errors.

## Usage Notes

There are no constraints on the use of the images and associated data. The data descriptors file lists all samples and assays, one for each section image, and also a “source” which for this material is the original embryo sample used by Kaufman in producing the histological sections. The data descriptor file lists the embryo code that Kaufman provided in his paper atlas and which can be used to identify the set of physical glass slides held by the University of Edinburgh on which each histological section can be found. In principle it is possible to obtain further images of the same or other sections in the series.

## Acknowledgements

The authors would like to acknowledge the critical contribution to the collection, collation and curation of these data by Liz Graham and Julie Moss and to the development of the data visualisation aspects by Nick Burton. All these are co-authors on the key citation for this work [Graham et al, 2015]. The authors would also like to thank the support of the Edina DataShare team, in particular Pauline Ward and Robin Rice. Finally this work has been funded as part of the MRC Funded core-funded Mouse Atlas Programme at the MRC Human Genetics Unit Edinburgh.

## Author contributions

RAB leads the Mouse Atlas programme that generated these datasets, designed the dataset submissions to the DataShare system and wrote the scripts providing the upload formats needed for batch ingest to DataShare. He also wrote the first draft of the manuscript. CA is the senior editor for the Mouse Atlas databases, co-wrote the manuscript and has performed much of the quality control on the datasets.

## Competing interests

The authors declare they have no competing interests in the publication of this data and manuscript.

## Figures (optional)

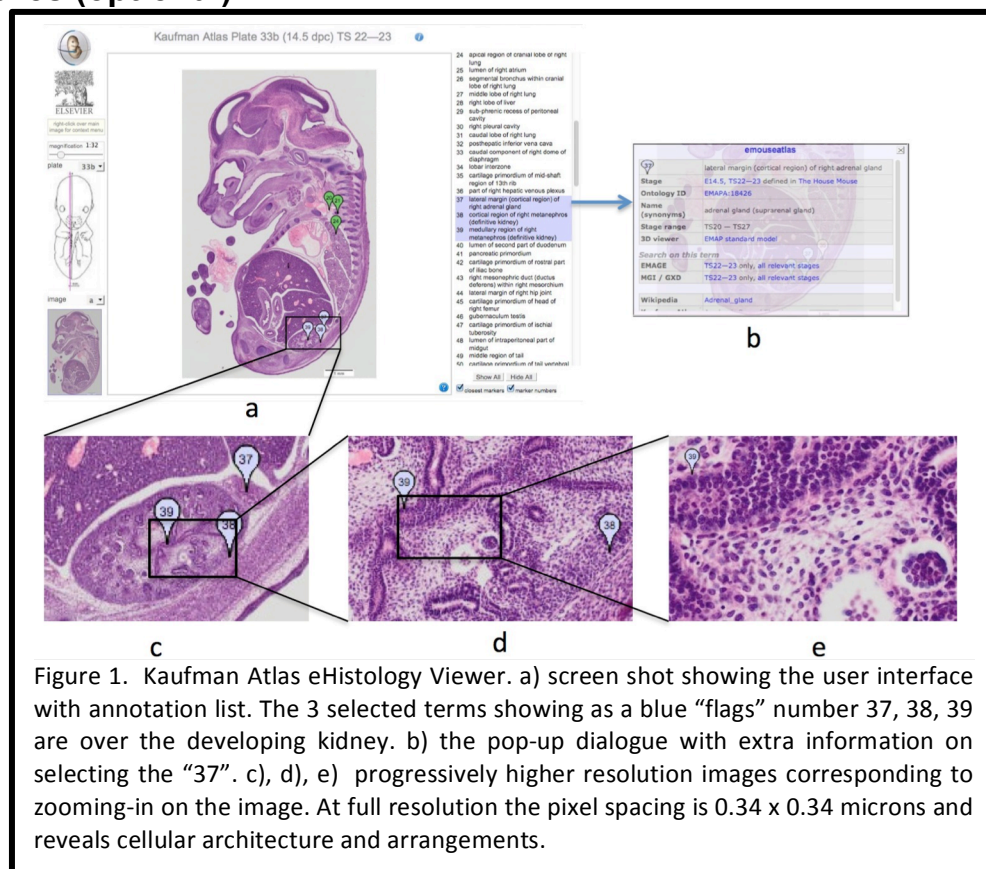


Figure 1. Kaufman Atlas eHistology Viewer. a) screen shot showing the user interface with annotation list. The 3 selected terms showing as a blue “flags” number 37, 38, 39 are over the developing kidney. b) the pop-up dialogue with extra information on selecting the “37”. c), d), e) progressively higher resolution images corresponding to zooming-in on the image. At full resolution the pixel spacing is 0.34 x 0.34 microns and reveals cellular architecture and arrangements.

## Figure Legends (optional)

Figure 1. Kaufman Atlas eHistology Viewer. a) screen shot showing the user interface with annotation list. The 3 selected terms showing as a blue “flags” number 37, 38, 39 are over the developing kidney. b) The pop-up dialog with extra information on selecting the “37”. c), d), e) progressively higher resolution images corresponding to zooming-in on the image. At full resolution the pixel spacing is 0.34 x 0.34 microns and reveals cellular architecture and arrangements.

## Tables

Table legends from above:

Table 1. Listing of the data files available with each dataset.

Table 2. DataShare Dublin Core elements used for the Kaufman datasets. The DCMI column provides the official Dublin Core term for the element, the Label is the heading for this data on the DataShare metadata listing.

Table 3. Complete list of data records for the Kaufman Atlas image set. The DOI resolves to a data set of image data, metadata and annotations, which can be downloaded individually or combined. Image volumes range up to about 2Gb with a total volume for the full set of compressed zip files 118Gb. The Position column gives an estimate of the relative distance through the embryo of the individual histology section. The values are between 0 (zero) and 1 (one) corresponding to the proportionate distance left-right (sagittal sections), cranial-caudal (transverse sections) and dorsal-ventral (coronal sections).

## References

1. Graham, E., Moss, J., Burton, N., Roochun, Y., Armit, C., Lorna Richardson, L. & Baldock, R. The atlas of mouse development eHistology resource. *Development* **142**, 1909-1911 (2015).
2. Richardson L, Moss J, Graham E, Burton N, Armit C, Baldock R (2015) Developing the eHistology Atlas. *Database (Oxford)* 2015. . **bav105** doi: 10.1093/database/bav105
3. Kaufman, MH. The Atlas of Mouse Development. *Elsevier Academic Press, Amsterdam, The Netherlands*. (1994).
4. Hayamizu, TF., Wicks, MN., Davidson, DR., Burger, A., Ringwald, M. & Baldock, RA. EMAP/EMAPA ontology of mouse developmental anatomy: 2013 update, *Journal of Biomedical Semantics*. **4:15**. (2013).
5. Hayamizu, TF., Baldock, RA. & Ringwald, M., Mouse anatomy ontologies: enhancements and tools for exploring and integrating biomedical data, *Mamm Genome*. **26**: 422–430 (2015).
6. Brown, SD. & Moore, MW. Towards an encyclopaedia of mammalian gene function: the International Mouse Phenotyping Consortium. *Dis Model Mech* **5**(3), 289–92 (2012).
7. Mohun, T., Adams, DJ., Baldock, R., Bhattacharya, S., Copp, AJ., Hemberger, M., Houart, C., Hurler, ME., Robertson, E., Smith, JC., Weaver, T. & Weninger, W. Deciphering the Mechanisms of Developmental Disorders (DMDD): a new programme for phenotyping embryonic lethal mice. *Disease Models & Mechanisms* **6**, 562-566 (2013).
8. Attwood, TK., Agit, B. & Ellis, ABM. Longevity of Biological Databases. *EMBnet.journal* **21**, e803 (2015). <http://dx.doi.org/10.14806/ej.21.0.803>
9. Rice, R. Edinburgh datashare – reflections from a data repository manager. *Bull. Assoc. Information Science and technology* **40(2)** 39-40 (2014). DOI: 10.1002/bult.2014.1720400212



10. Richardson, L., Venkataraman, S., Stevenson, P., Yang, Y., Moss, J., Graham, L., Burton, N., Hill, B., Rao, J., Baldock, RA. & Armit, C. EMAGE mouse embryo spatial gene expression database: (2014 update). *Nucleic Acids Res.* **42(1)** D835-44 (2014). doi: 10.1093/nar/gkt1155

## Data Citations

Bibliographic information for the data records described in the manuscript.

1. Graham, E., Moss, J., Burton, N., Roochun, Y., Armit, C., Lorna Richardson, L. & Baldock, R. *DataShare* <http://dx.doi.org/10.7488/ds/393> to <http://dx.doi.org/10.7488/ds/1305>